

DNA: Structure, Function, and Replication

1. Introduction

DNA, or **deoxyribonucleic acid**, is the fundamental **molecule of heredity** in almost all living organisms. It serves as the blueprint for building and regulating the functions of cells and, ultimately, entire organisms. Every cell (except some like mature red blood cells) carries DNA, which stores genetic information in a **precise, code-like sequence** of four chemical bases. The discovery of DNA's double-helical structure by **James Watson and Francis Crick in 1953**, supported by **Rosalind Franklin's X-ray diffraction images**, revolutionized biology and paved the way for molecular genetics, biotechnology, and genome sequencing.

2. Structure of DNA

DNA's structure is a masterpiece of molecular architecture, perfectly designed for **stability, replication, and information storage**.

2.1 Double Helix Model

- **Two Antiparallel Strands:** The two DNA strands run in opposite directions — one from **5' to 3'** and the other from **3' to 5'** — which is essential for accurate replication.
- **Nucleotides:** The building blocks of DNA. Each nucleotide is composed of:
 - **Deoxyribose sugar** (a 5-carbon sugar lacking an oxygen atom at the 2' position).
 - **Phosphate group** (forms the backbone through phosphodiester bonds).
 - **Nitrogenous base:** adenine (**A**), thymine (**T**), guanine (**G**), cytosine (**C**).
- **Base Pairing Rules:**
 - **Adenine (A)** pairs with **Thymine (T)** via **two hydrogen bonds**.
 - **Guanine (G)** pairs with **Cytosine (C)** via **three hydrogen bonds**.
 - Complementary base pairing ensures **replication accuracy** and **structural stability**.

2.2 Helical Configuration

- DNA is typically in the **B-form** (right-handed helix with ~10.5 base pairs per turn).
- Variants like **A-DNA** (shorter, dehydrated form) and **Z-DNA** (left-handed, transient form) can appear under special physiological conditions.

3. Organization of DNA in Cells

The **DNA molecule** is extraordinarily long relative to the size of the cell. In a single human cell, if fully stretched, the DNA from all chromosomes would measure about **2 meters in length**, yet it must fit inside a cell nucleus that is only about **6 micrometers** in diameter. This creates a remarkable challenge: the genetic material needs to be **highly compacted** to fit into this microscopic space, while at the same time remaining **accessible** for critical processes like **gene expression, DNA replication, and repair**.

Cells achieve this through a **hierarchical packaging system** involving proteins, structural organization, and dynamic changes in chromatin state.

3.1 Nucleosomes and Chromatin

Histone Proteins – The First Level of DNA Packaging

The fundamental unit of DNA packaging is the **nucleosome**. Here, **DNA is wrapped around an octamer of histone proteins** — two copies each of **H2A, H2B, H3, and H4**. About **147 base pairs** of DNA coil around this histone core, forming a structure often described as “**beads on a string**” when observed under an electron microscope.

This arrangement not only compacts the DNA but also **organizes it in a predictable, repeatable pattern** that can be modified for gene regulation.

Linker DNA and Histone H1 – Higher-Order Compaction

Between nucleosomes lies **linker DNA**, a short stretch of DNA not wrapped around histones. The **linker histone (H1)** binds to this region and helps **stabilize the angle** at which the DNA enters and exits each nucleosome. This action facilitates the folding of the “beads-on-a-string” structure into a **more compact 30-nanometer fiber**, an intermediate stage of chromatin condensation.

Chromatin States – Functional Organization

Chromatin is **not static**; it exists in different compaction states, which influence **gene activity**:

- **Euchromatin:**
 - Loosely packed and **gene-rich**.

- Accessible to **RNA polymerase** and transcription factors.
- Associated with **active gene expression**.
- Found more toward the **interior of the nucleus**.
- **Heterochromatin:**
 - Densely packed and **gene-poor**.
 - Transcriptionally silent, often containing **repetitive sequences**.
 - Found at **nuclear peripheries** or around **centromeres and telomeres**.
 - Plays a role in **structural stability** and protecting chromosome ends.

The cell can dynamically shift between these states by **histone modification, DNA methylation, and chromatin remodeling**, allowing for responsive control over gene activity.

3.2 Chromosomes: The Highest Level of DNA Organization

When a cell is not dividing, its DNA exists mainly in the **chromatin form** — a relaxed yet organized structure that allows access for transcription. However, during **cell division** (mitosis and meiosis), chromatin **condenses** into the more **distinct and tightly packed structures called chromosomes**. This condensation ensures **accurate segregation of genetic material** into daughter cells.

- **Human Chromosome Number:** Human somatic (body) cells contain **46 chromosomes** arranged in **23 pairs**:
 - **22 pairs of autosomes** (non-sex chromosomes)
 - **1 pair of sex chromosomes** (XX in females, XY in males)
- **Chromosome Inheritance:** Each pair consists of one chromosome from the **mother** and one from the **father**. This arrangement ensures genetic diversity and maintains a constant chromosome number across generations.
- **Structural Features:**
 - **Centromere:** Constriction point dividing the chromosome into a short arm (**p**) and a long arm (**q**), crucial for proper segregation.
 - **Telomeres:** Protective caps of repetitive DNA sequences at chromosome ends that prevent degradation and fusion.

- **Origins of Replication:** Specific sites where DNA replication begins during S-phase.

4. Functions of DNA

DNA is far more than a static chemical structure; it is the **master blueprint of life**. It not only stores biological information but also orchestrates the activities of cells, ensures continuity across generations, and fuels the process of evolution.

4.1 Information Storage

DNA holds genetic instructions in the form of a **sequence of nucleotides** — adenine (A), thymine (T), cytosine (C), and guanine (G). These sequences are arranged in **triplets called codons**, each of which specifies a particular amino acid.

- This genetic code is **universal** (with minor exceptions in some organelles and microorganisms), meaning the same codon specifies the same amino acid across almost all life forms.
- The linear sequence of bases in DNA can be compared to letters in a long instruction manual, where the order determines the structure and function of proteins, which in turn dictate cellular processes.

4.2 Transmission of Heredity

DNA ensures that organisms resemble their parents because it is faithfully **replicated** before cell division.

- In **mitosis**, this ensures that daughter cells are genetically identical to the parent cell.
- In **meiosis**, it ensures that gametes carry half the genetic material, allowing traits to be inherited when fertilization restores the full set of chromosomes.
- The accuracy of DNA replication, aided by proofreading enzymes like **DNA polymerase**, minimizes errors, though occasional mutations provide raw material for evolution.

4.3 Guiding Protein Synthesis

Proteins are the workhorses of the cell, and DNA provides the instructions for making them through two major steps:

1. **Transcription** – A gene's DNA sequence is copied into messenger RNA (mRNA) by RNA polymerase.

2. **Translation** – Ribosomes read the mRNA codons and assemble the corresponding amino acids into a polypeptide chain.
- This process follows the **central dogma of molecular biology: DNA → RNA → Protein**.
 - Some genes code for RNA molecules (like rRNA and tRNA) that are directly functional without being translated into proteins.

4.4 Regulatory Roles

Not all DNA directly codes for proteins; a large proportion is **non-coding DNA** that plays a vital role in regulating gene expression.

- **Promoters, enhancers, and silencers** are DNA sequences that influence when and how strongly a gene is transcribed.
- **Epigenetic modifications** (like DNA methylation and histone acetylation) can turn genes on or off without altering the underlying sequence.
- Regulatory DNA ensures that **the same genetic material can produce very different cell types**, depending on which genes are active — for example, a neuron vs. a muscle cell.

5. DNA Replication in Eukaryotes

DNA replication is the essential process by which a eukaryotic cell makes an exact copy of its DNA before it divides. This ensures that **each daughter cell receives a complete and identical set of genetic instructions**, preserving the organism's genetic identity across generations of cells.

Eukaryotic DNA replication follows a **semi-conservative model**—meaning that each new DNA molecule contains **one original parental strand** and **one newly made daughter strand**. This was first suggested by Watson and Crick and later confirmed by the famous **Meselson–Stahl experiment (1958)**.

Because eukaryotic genomes are **large, linear, and stored within a nucleus**, replication is **slower and more complex** than in prokaryotes. Many **specialized enzymes and regulatory proteins** work in a highly coordinated sequence to ensure accuracy. The process occurs in **three main stages: Initiation, Elongation, and Termination**.

5.1 Initiation

The **initiation phase** sets up the replication machinery at **specific DNA sites** called **origins of replication (Ori)**. In eukaryotes, each chromosome has **thousands of such origins** to ensure the entire genome can be copied quickly.

Step-by-Step Events in Initiation:

1. Origin of Replication (Ori)

- Special DNA sequences act as **starting points**.
- Recognized by a protein complex called the **Origin Recognition Complex (ORC)**, which acts like a “landing pad” for other replication proteins.
- Multiple origins work simultaneously across the genome.

2. Unwinding the DNA: DNA Helicase

- Helicase attaches to the DNA at the origin and **unzips the double helix** by breaking **hydrogen bonds** between complementary base pairs.
- This forms a **replication fork**—a Y-shaped structure where new DNA strands will be built.

3. Stabilizing the Open DNA: Replication Protein A (RPA)

- In eukaryotes, RPA binds to the single-stranded DNA to **prevent re-annealing** and protect the DNA from being degraded by enzymes.
- This keeps the DNA **single-stranded and accessible** for copying.

4. Laying the Starting Blocks: Primase

- Primase is a special **RNA polymerase** that builds a short **RNA primer** (~10–15 nucleotides in eukaryotes).
- DNA polymerases **cannot start on their own**; they can only add nucleotides to an existing strand, so primers act as the **starting platform** for synthesis.

5.2 Elongation

Once the DNA is open and primed, the cell begins **copying the DNA strand by strand**.

DNA Polymerases in Eukaryotes:

- DNA polymerase ϵ (epsilon) → Synthesizes the **leading strand**.
- DNA polymerase δ (delta) → Synthesizes the **lagging strand**.

Leading Strand Synthesis

- Made **continuously** in the same direction as the replication fork moves.

- Needs only **one primer** for the entire length.

Lagging Strand Synthesis

- Made **discontinuously**, opposite to fork movement, in short segments called **Okazaki fragments**.
- In eukaryotes, these fragments are about **100–200 nucleotides long**.
- Each fragment starts with a primer.

Processing and Joining

1. **Primer Removal** – In eukaryotes, **RNase H** and DNA polymerase δ remove RNA primers and replace them with DNA.
2. **Sealing the Gaps** – **DNA ligase** joins the Okazaki fragments by forming **phosphodiester bonds**, creating one continuous strand.

Proofreading

- DNA polymerases ϵ and δ have **3'→5' exonuclease activity**—like a built-in proofreading tool.
- If a wrong nucleotide is inserted, the polymerase **backs up**, removes the error, and replaces it correctly.
- This gives replication a **very low error rate** (~1 error per 10 million bases, further reduced by repair systems).

5.3 Termination

The replication process ends once the **entire chromosome has been copied**.

In Eukaryotes:

- **Replication forks meet** from opposite directions, ensuring no gaps remain.
- The major challenge: **replicating the ends of chromosomes**.

Telomeres and Telomerase

- **Telomeres** are repetitive, non-coding sequences at the ends of chromosomes (e.g., TTAGGG in humans).
- DNA polymerases cannot fully replicate the very ends, leading to gradual shortening in most cells.
- **Telomerase** solves this in germ cells, stem cells, and certain white blood cells:

- It's an **enzyme–RNA complex** that uses its own RNA template to **extend the parental strand**.
- This allows the lagging strand to be fully completed.
- In most **somatic cells**, telomerase is inactive, so telomeres shorten over time, contributing to **cellular aging**.

Key Points Summary

- **Semi-conservative:** One old + one new strand per DNA molecule.
- **Bidirectional replication:** Forks move in opposite directions from the origin.
- **Enzyme coordination:** Helicase unwinds, SSBs stabilize, primase initiates, DNA polymerase elongates, ligase seals, topoisomerase relieves supercoiling.
- **Eukaryotic complexity:** Multiple origins, chromatin remodeling, telomere maintenance.

6. Replication Fidelity and Mutations

DNA replication is an extraordinarily precise process, vital for the accurate transmission of genetic information from one cell generation to the next. However, this precision is not absolute—errors can occur, and how cells detect, correct, or tolerate them determines both genetic stability and the potential for evolutionary change.

Proofreading and Repair

During replication, **DNA polymerases**—the enzymes that synthesize new DNA strands—possess an inherent **proofreading function**. As they add nucleotides, they continuously check whether the newly incorporated base correctly pairs with the template strand.

If a mismatch is detected, the enzyme removes the incorrect nucleotide through its **3'→5' exonuclease activity** and replaces it with the correct one before proceeding. Beyond proofreading, cells possess a range of **DNA repair mechanisms**:

- **Mismatch repair** identifies and corrects base-pairing errors that escape proofreading, using the original template strand as a guide.
- **Nucleotide excision repair** removes damaged stretches of DNA, such as those caused by UV-induced thymine dimers, and replaces them with correctly paired nucleotides.
- **Base excision repair** targets individual chemically altered bases, excising and replacing them with the correct forms.

These systems maintain an extraordinarily low error rate—often as few as **one mistake per billion nucleotides replicated**—ensuring genetic continuity.

Mutations

When errors remain uncorrected, they result in **mutations**, permanent changes to the DNA sequence. Mutations can occur in **coding regions**, potentially altering protein structure and function, or in **regulatory sequences**, affecting when and how genes are expressed.

Some mutations are **harmful**, leading to genetic disorders such as:

- **Cystic fibrosis** – caused by mutations in the CFTR gene affecting chloride ion transport.
- **Sickle cell anemia** – due to a single nucleotide change in the β -globin gene, altering hemoglobin structure.

External Factors Influencing Mutation Rates

While replication errors are one source of mutations, many arise from **external mutagens**—environmental agents that damage DNA:

- **Ultraviolet (UV) light** induces covalent bonds between adjacent thymine bases, distorting the DNA helix.
- **Ionizing radiation** (e.g., X-rays, gamma rays) can cause double-strand breaks in DNA.
- **Chemical mutagens** such as benzene derivatives or alkylating agents can modify bases, leading to mispairing during replication. Cells respond to such insults by activating DNA damage response pathways, halting the cell cycle to allow repair or, if damage is irreparable, triggering **apoptosis** (programmed cell death) to prevent the propagation of defective DNA.